Enzymology of base excision repair in the hyperthermophilic archaeon Pyrobaculum aerophilum

Sartori, Alessandro A ; Jiricny, J

Abstract: DNA of all living organisms is constantly modified by exogenous and endogenous reagents. The mutagenic threat of modifications such as methylation, oxidation, and hydrolytic deamination of DNA bases is counteracted by base excision repair (BER). This process is initiated by the action of one of several DNA glycosylases, which removes the aberrant base and thus initiates a cascade of events that involves scission of the DNA backbone, removal of the baseless sugar-phosphate residue, filling in of the resulting single nucleotide gap, and ligation of the remaining nick. We were interested to find out how the BER process functions in hyperthermophiles, organisms growing at temperatures around 100 degrees C, where the rates of these spontaneous reactions are greatly accelerated. In our previous studies, we could show that the crenarchaeon Pyrobaculum aerophilum has at least three uracil-DNA glycosylases, Pa-UDGa, Pa-UDGb, and Pa-MIG, that can initiate the BER process by catalyzing the removal of uracil residues arising through the spontaneous deamination of cytosines. We now report that the genome of P. aerophilum encodes also the remaining functions necessary for BER and show that a system consisting of four P. aerophilum encoded enzymes, Pa-UDGb, AP endonuclease IV, DNA polymerase B2, and DNA ligase, can efficiently repair a G.U mispair in an oligonucleotide substrate to a G.C pair. Interestingly, the efficiency of the in vitro repair reaction was stimulated by Pa-PCNA1, the processivity clamp of DNA polymerases.

DOI: https://doi.org/10.1074/jbc.M302397200

DNA Replication Repair and Recombination:

Enzymology of base excision repair in the hyperthermophilic archaeon pyrobaculum aerophilum

Alessandro A. Sartori and Josef Jiricny

J. Biol. Chem. published online April 30, 2003

JBC Papers in Press. Published on April 30, 2003 as Manuscript M302397200

Classification: Enzymology

Enzymology of Base Excision Repair in the Hyperthermophilic Archaeon *Pyrobaculum aerophilum*.

Alessandro A. Sartori and Josef Jiricny*

Institute of Molecular Cancer Research

University of Zürich

August Forel-Strasse 7, CH-8008 Zürich, Switzerland

Running TitleBase Excision Repair in *Pyrobaculum aerophilum*

*Communicating author

Tel.: +41-1-634 8910

Fax.: +41-1-634 8904

E-mail: jiricny@imr.unizh.ch

SUMMARY

DNA of all living organisms is constantly modified by exogenous and endogenous reagents. The mutagenic threat of modifications such as methylation, oxidation and hydrolytic deamination of DNA bases is counteracted by base excision repair (BER). This process is initiated by the action of one of several DNA glycosylases, which removes the aberrant base and thus initiates a cascade of events that involves scission of the DNA backbone, removal of the baseless sugar-phosphate residue, filling-in of the resulting single nucleotide gap and ligation of the remaining nick. We were interested to find out how the BER process functions in hyperthermophiles, organisms growing at temperatures around 100ÚC where the rates of these spontaneous reactions are greatly accelerated. In our previous studies we could show that the crenarchaeon *Pyrobaculum aerophilum* has at least three uracil-DNA glycosylases, *Pa*-UDGa, *Pa*-UDGb and *Pa*-MIG that can initiate the BER process by catalyzing the removal of uracil residues arising through the spontaneous deamination of cytosines. We now report that the genome of *P. aerophilum* encodes also the remaining functions necessary for BER and show that a system consisting of four *P.aerophilum* encoded enzymes, *Pa*-UDGb, AP endonuclease IV, DNA polymerase B2 and DNA ligase, can efficiently repair a G•U mispair in an oligonucleotide substrate to a G•C pair. Interestingly, the efficiency of the *in vitro* repair reaction was stimulated by *Pa*-PCNA1, the processivity clamp of DNA polymerases.

INTRODUCTION

Most hyperthermophiles, organisms living at temperatures of around 100°C, belong to *Archaea* (1), the closest prokaryotic relatives of eukaryotes (2). Two key pathways of DNA metabolism, transcription and replication, are highly conserved among *Archaea* and eukaryotes (3, 4), and we were interested to learn whether these similarities extended also to the third domain of DNA metabolism, namely, DNA repair. We chose to study the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*, the genomic sequence of which has recently become available (5).

DNA repair processes can be classified into three major categories: damage reversal (DR), recombination repair (RR) and excision repair; the latter can be further subdivided into three biochemical pathways: nucleotide excision repair (NER), mismatch repair (MMR) and base excision repair (BER). Sequence similarity searches for DNA repair genes in *P. aerophilum* revealed that the organism apparently lacks key representatives of several of these pathways. Thus, we found only a single representative of the DR class, O^6-alkylguanine DNA alkyltransferase, which is present in most *Archaea* (6), and two copies of the RecA/RAD51 recombinase homologue RadA. It is not known whether the latter genes encode functional polypeptides, as the genome of this organism appears to carry little evidence of recombination events (5). This implies that RR may be rather infrequent, as shown for *Sulfolobus acidocaldarius* (7), a close relative of *P. aerophilum*. Only two putative homologues of mammalian NER enzymes, the DNA helicases XPB and XPD (5), could be found, and it is presently unclear whether these proteins are involved in DNA repair or transcription, as components of the transcription factor TFIIH (8). Like other *Archaea*, the genome of *P. aerophilum* carries no homologues of the MMR genes *mutS* and *mutL*.

Although this does not rule out the existence of a MMR system, the mutation frequency in *P. aerophilum* is rather high, especially in mononucleotide repeats, which is consistent with the absence of a MMR pathway (5).

The remaining pathway, BER, is responsible for the removal of the largest fraction of DNA damage, which is mainly associated with the modification or loss of DNA bases through hydrolysis, oxidation or methylation. Spontaneous hydrolytic reactions lead to deamination of cytosine, 5-methylcytosine and adenine in DNA to give rise to uracil, thymine and hypoxanthine, respectively, or to the loss of purines. Given that the rates of most chemical reactions double with each 10ÚC increase in temperature, it is to be expected that the DNA of hyperthermophiles, which live at temperatures around 100ÚC, would be modified to a much greater extent than that of organisms living in ambient environments. Correspondingly, BER would be expected to assume greater importance for the survival of the former organisms. Our earlier studies appear to substantiate this hypothesis, as the 2.2Mb genome of *P. aerophilum* encodes at least three different uracil-DNA glycosylases, *Pa*-MIG, *Pa*-UDGa and *Pa*-UDGb (9-11), which initiate the repair of cytosine deamination by excising uracil from DNA [reviewed in (12-14)].

The glycosylase-mediated removal of damaged or modified bases gives rise to abasic (apyrimidinic or apurinic) sites (AP-sites) in the DNA, which are non-coding and therefore highly mutagenic (15). Their repair involves in the first instance an AP endonuclease-catalyzed incision of the sugar-phosphate backbone at the 5' side of the AP-site, which gives rise to a single-strand break where the upstream fragment is terminated with a free 3'-hydroxyl group, while the downstream one carries a baseless deoxyribose-phosphate (dRP) residue at its 5'-terminus. In mammalian cells, the subsequent processing of the incised AP-

sites can be accomplished by one of two distinct pathways: short-patch BER, the preferred mechanism that results in the replacement of a single nucleotide residue, and long-patch BER, where the repair tracts are 2-6 nucleotides long (16). In the short-patch pathway, DNA polymerase β (pol-β) extends the upstream fragment by a single nucleotide, and concomitantly removes the dRP moiety by β-elimination (17). The DNA ligase III/XRCC1 complex (18, 19) then seals the remaining nick. In long-patch BER, the upstream primer is thought to be extended by pol-β (or pol-δ) by 2-6 nucleotides, which results in the displacement of the dRP residue as part of a 'flap' oligonucleotide (20). This short overhang is then excised by the flap-endonuclease (FEN1), and the resulting nick is sealed by DNA ligase I. Proliferating cell nuclear antigen (PCNA) may also be involved in this process, as it was found to enhance the pol-β-dependent long patch BER by stimulating the activity of FEN1 and to interact with DNA ligase I (21, 22). The long-patch pathway may be important for the repair of reduced or oxidized AP-sites, in which the modified dRP residues are resistant to pol-β-catalyzed β-elimination reactions (21).

We set out to explore the BER system of *P. aerophilum*, the genome of which carries, in addition to the above-mentioned DNA glycosylase genes, also genes encoding orthologues of downstream-acting BER proteins: one putative AP endonuclease, three putative DNA polymerases of the B family, one putative DNA ligase and PCNA1, which was already shown to interact with FEN1, UDGa and PolB3 from *P. aerophilum* (23). We expressed these proteins in *E. coli*, and analyzed their ability to catalyze the repair of a G•U mispair in a double-stranded oligonucleotide substrate. In the following sections we show that efficient repair of this substrate can be accomplished by the *P. aerophilum* proteins UDGb, AP endonuclease IV, DNA polymerase B2 and DNA ligase. The putative role of *Pa*-PCNA1 in

this process is also discussed.

EXPERIMENTAL PROCEDURES

Reagents and Oligonucleotides- All oligonucleotides were synthesized by Microsynth (Balgach, Switzerland) and were purified by PAGE. Restriction enzymes, T4 DNA Ligase, T4 DNA Polymerase, T4 Polynucleotide Kinase and *E. coli* UDG were supplied by New England BioLabs (Beverly, MA). *Pfu* DNA Polymerase and *Taq* DNA Polymerase were from Stratagene and Qiagen, respectively. All other chemicals and reagents were purchased from Sigma, Roche Molecular Biochemicals, Amresco, Epicentre Technologies or Merck, and were of analytical grade purity.

P. aerophilum Extracts and Purified Proteins- P. aerophilum cultures were grown in the laboratory of Jeffrey Miller (University of California, Los Angeles, CA) and cell-free extracts were prepared by Mahmud Shivji as described previously (24). The cell-free extract was supplemented with 1 mM PMSF and 1 x Complete", EDTA-free (Boehringer Mannheim) protease-inhibitor cocktail and dialyzed overnight at 4°C against 2 liters of buffer containing 25mM sodium phosphate pH 8.0, 50mM NaCl, 10 % glycerol, 0.5mM EDTA and 2mM DTT. The protein concentration was estimated to be 4 mg/ml as determined by the method of Bradford (25). Aliquots were stored at –80°C. The recombinant *P. aerophilum* uracil-DNA glycosylases *Pa*-UDGa and *Pa*-UDGb were expressed and purified as described previously (10,11). The purified recombinant wild-type human DNA polymerase-β was a kind gift of Samuel H. Wilson (Laboratory of Structural Biology, NIEHS, NC).

Bacterial Strains and Expression Plasmids- The *E. coli* strain DH5α was used in all cloning experiments and for plasmid amplifications, and the strain BL21 (DE3) was used for protein expression (26). The plasmid pET28c(+) (Novagen) was used for bacterial expression

of N-terminal His_6-tagged proteins. The plasmids pQE30-*PaPolB3* and pQE30-*PaPCNA1* for bacterial overexpression of full-length *P. aerophilum* proteins DNA polymerase B3 (*Pa*-PolB3; PAE2109: 785 amino acids, 89.5 kDa) and PCNA1 (*Pa*-PCNA1; PAE3038: 249 amino acids, 28 kDa) containing an N-terminal His_6-tag were a kind gift of Hanjing Yang (University of California, Los Angeles, CA). The gene encoding *Pa*-PCNA1 was subloned from pQE30 into pET28c(+) vector using *Nde*I and *Hin*dIII restriction sites.

Computational Analyses and Cloning of the P. aerophilum AP endonuclease IV, DNA Ligase and DNA Polymerase B2 genes (PaEndoIV, PaDNA-Ligase and PaPolB3)- Theopen reading frames PAE3257 (275 amino acids, 31.1 kDa), PAE833 (589 amino acids, 65.1 kDa), and PAE1113 (553 amino acids, 61.5 kDa) encoding the putative proteins *Pa*-EndoIV, *Pa*-DNA-Ligase, and *Pa*-PolB2, respectively, were identified in the complete genome sequence of *P. aerophilum* (5). Analyses of the *P. aerophilum* genome were performed using the Genetics Computer Group program package, Version 10, 1999 (27); for other database searches we used the BLAST and ENTREZ services available on the National Center for Biotechnology Information (NCBI) Web site (www.ncbi.nlm.nih.gov). The respective ORFs were amplified by PCR from *P. aerophilum* genomic DNA using the following sense (s) and antisense (as) primers for subsequent cloning into pET28c(+) vector (restriction sites are in bold letters): EndoIV-s: 5'-GGATC**GCTAGC**ATGGCAAAGGTATATCTGGGGCCTGC-3' and EndoIV-as: 5'-GTAC**GGATCC**CTATGCCAAGTTTACGCCGACTTGC-3'; PolB2-s: 5'-GGATC**CATATG**TTCGTAATTGGGGCCAGGCCG-3' and PolB2-as: 5'-GTAC**CT-CGAG**TCATATAAGCCTTAAGGCGCGTAACTTCC-3'; DNA-Ligase-s: 5'-GGATC**C-ATATG**GGTAGTATATACGTGCAGTTTGGGGAG and DNA-Ligase-as: 5'-GTAC**G-GATCC**CTACTCGGGCTGAACCACTTTTTTCTGC.

Expression of the recombinant Pa-EndoIV, Pa-DNA-Ligase, Pa-PolB2, Pa-PolB3 and Pa-PCNA1 proteins- The His_6-tagged fusion proteins were produced by transforming *E. coli* BL21 (DE3) cells with the respective expression constructs (pET28-*PaEndoIV,* pET28-*PaDNA-Ligase*, and pET28–*PaPolB2*). Following incubation overnight at 30°C on selective LB agar plates containing 50 µg/ml kanamycin (LB-kan) and 2% D-glucose, single colonies were picked and grown overnight at 30°C in 20 ml LB-kan medium supplemented with 2% D-glucose. The saturated cultures were diluted 1:100 in 1 liter of LB-kan medium and grown at 37°C until the OD_{600} reached 0.5-0.8. The cultures were cooled down to 30°C and the overexpression of the recombinant proteins was induced with 250µM IPTG. The cells were then grown overnight at 30°C and pelleted by centrifugation (Sorvall SS34 rotor, 4000 rpm, 30 min) at 4°C.

The recombinant *Pa*-PolB3 protein (PAE2109; 785 amino acids, 89.5 kDa) was overexpressed using the above procedure, except that the BL21 (DE3) cells were co-transformed with two plasmids at once: the expression construct pQE30-*PaPolB3* bearing the gene for ampicillin resistance, and pREP4, bearing the gene for kanamycin resistance. Therefore, both LB agar plates and LB medium had to be supplemented with ampicillin and kanamycin. The pREP4 plasmid constitutively expresses the Lac repressor protein encoded by the lacI gene to reduce the basal level of expression (Qiagen).

The recombinant *Pa*-PCNA1 protein (PAE3038; 249 amino acids, 28kDa) was produced by transforming *E. coli* BL21 (DE3) cells with the pET28-*PaPCNA1* expression construct. Following incubation overnight at 30°C on selective LB agar plates containing 50 µg/ml kanamycin (LB-kan) and 2% D-glucose, a single colony was picked and grown overnight at 30°C in 20 ml LB-kan medium supplemented with 2% D-glucose. The saturated culture was

diluted 1:100 in 1 liter of LB-kan medium and grown at 37°C until the OD_{600} reached 1.0. The culture was cooled to room temperature and the overexpression of the recombinant protein was induced with 1mM IPTG. After growing the cells overnight at 30°C, they were pelleted by centrifugation as above.

Purification of Pa-EndoIV and Pa-DNA-Ligase- The cell pellets were resuspended in 3 ml/g of ice-cold sonication buffer (50mM sodium phosphate pH 8.0, 300mM NaCl, 10% glycerol, 1mM imidazole, 0.25% Tween-20, 10mM β-mercaptoethanol and 1mM PMSF) and the cells were lysed by sonication with 25 x 10 s bursts on ice. The cell debris was removed by centrifugation at 15000 rpm for 30 min at 4°C in a Sorvall SS34 rotor. The soluble fraction of the whole cell-free extract was heat-treated for 30 min at 70°C and the precipitated *E.coli* proteins were removed by centrifugation as above. The supernatant, containing only thermo-stable proteins, was incubated with gentle shaking for 1 h at 4°C with 1 ml of Ni-NTA-agarose (Qiagen), pre-equilibrated in sonication buffer. The suspension was then packed into a disposable column and unbound proteins were eluted with sonication buffer containing increasing concentrations of imidazole (1 x 20 ml 5mM imidazole and 3 x 10 ml 20mM imidazole). The His_6-tagged protein was eluted with 3 x 1ml sonication buffer containing 250mM imidazole. These fractions were pooled and diluted to 100mM NaCl in Buffer A (25mM sodium phosphate pH 8.0, 10% glycerol and 5mM β-mercaptoethanol) and loaded onto a 1 ml Hi-Trap Q Sepharose column (Pharmacia). The flowthrough containing the recombinant protein was dialyzed overnight at 4°C against 1 liter of Buffer A supplemented with 25mM NaCl and loaded onto a 1 ml Hi-Trap Heparin column (Pharmacia). The column was extensively washed with Buffer A and the bound proteins were eluted with a 5 ml linear gradient of 25-700mM NaCl. The nearly-homogenous *Pa*-EndoIV

and *Pa*-DNA-Ligase eluted at around 300mM NaCl. The peak fractions were pooled and dialyzed overnight at 4°C against 1 liter of Buffer A containing 50mM NaCl.

Purification of Pa-PolB2- The procedure, including Ni-NTA purification, was as above, except that the heat-treatment had to be omitted, as *Pa*-PolB2 was found to be less stable at higher temperatures. The 250mM imidazole fractions containing *Pa*-PolB2 were pooled and dialyzed overnight at 4°C against 1 liter of Buffer A supplemented with 50mM NaCl. The sample was loaded onto a 1 ml Mono-S FPLC column (Pharmacia) and the column was extensively washed with Buffer A. The proteins were then eluted with a 30-ml linear gradient of 50-700mM NaCl. The nearly-homogenous *Pa*-PolB2 protein eluted as a major peak in fractions containing 350-450mM NaCl. These fractions were pooled and dialyzed overnight at 4°C against 1 liter of Buffer A containing 50mM NaCl.

Purification of Pa-PolB3- Pa-PolB3 was purified similarly to *Pa*-EndoIV and *Pa*-DNA-Ligase, except that the heat-treatment was done at 80°C, as the protein exhibited the highest thermo-stability of all the proteins tested. After Ni-NTA purification, the 250mM imidazole fractions containing > 95% pure *Pa*-PolB3 were pooled and dialyzed overnight.

Purification of Pa-PCNA1- The cell pellet was resuspended in 3 ml/g of ice-cold sonication buffer (50mM Tris pH 7.5, 300mM NaCl, 10 % glycerol, 1mM imidazole, 0.25 % Tween-20, 10mM β-mercaptoethanol and 1mM PMSF) and the cells were lysed by sonication on ice. The cell debris was removed by centrifugation at 18000 rpm for 45 min at 4°C. The supernatant (soluble fraction) was incubated with gentle shaking for 30 min at 4°C with 1 ml of Ni-NTA-agarose (Qiagen). The suspension was then packed into a disposable column and unbound proteins were eluted with sonication buffer containing increasing concentrations of imidazole (1 x 20 ml 10mM imidazole and 3 x 10 ml 50mM imidazole).

The His_6-tagged *Pa*-PCNA1 protein was eluted with 3 x 1ml sonication buffer containing 500mM imidazole. These fractions were pooled and dialyzed overnight at 4°C against 1 liter of ion-exchange buffer (50mM Tris.HCl pH 7.5, 10mM NaCl, 10 % glycerol, 5mM β-mercaptoethanol) and loaded onto a Resource Q FPLC column (Pharmacia). The column was extensively washed with buffer and the bound proteins were eluted with a linear gradient of 10-700mM NaCl. The nearly-homogenous protein eluted at around 300mM NaCl. The peak fractions were pooled and dialyzed overnight at 4°C against 1 liter of buffer containing 20mM HEPES pH 7.5, 100mM NaCl, 10% glycerol and 5mM β-mercaptoethanol.

The concentration of the recombinant proteins, as measured by the method of Bradford (25) and verified by SDS-PAGE analysis, were as follows: *Pa*-EndoIV H 150 µg/ml, *Pa*-DNA-Ligase H 500 µg/ml, *Pa*-PolB2 H 300 µg/ml, *Pa*-PolB3 H 30 µg/ml and *Pa*-PCNA1 H 7.72 mg/ml. All proteins were stored in aliquots at 80°C.

Base excision repair assay using P. aerophilum whole cell-free extract (WCE)- The fluorescently-labeled substrates, containing either a single-nucleotide gap (1nt-gap) or a G•U mismatch at a defined position, were prepared as follows: 1-nt gap, the labeled 23-mer-F and 36-mer oligos were annealed with the complementary 60-mer G oligo; G•U, the labeled 60-mer U-F was annealed with the 60-mer G oligo according to the protocol described previously (10). The oligonucleotide sequences are listed in Table I and the substrates are schematically shown in Figure 1A.

All repair reactions (20 µl) contained 50mM Tris pH 8.0, 0.2mM EDTA, 1mM DTT, 1mM NAD, 50mM NaCl, 5 pmol of unlabeled single-stranded 50-mer oligo (50-mer G) and 1 pmol of labeled substrates. Essential cofactors, such as 5mM $MgCl_2$, 20µM of each dNTPs and 2mM ATP, were added as indicated in the figure. The reaction was initiated by the

addition of 20 µg of *P. aerophilum* WCE and allowed to proceed for 30 min at 60°C. In the cases indicated, repair was completed by the addition of T4 DNA Ligase, as the ligation step in the WCE was inefficient. After further incubation for 1 h at 30°C, reactions were terminated by the addition of 10 µl of stop solution (500 µg/ml Proteinase K, 5mM EDTA, 0.5 % SDS) and incubated for a further 30 min at 37°C. The DNA samples were then precipitated and the products analyzed as described previously (10).

AP endonuclease assays- DNA substrate oligos containing a single AP-site were prepared by incubating the labeled G•U* substrate with *E. coli* UDG for 20 min at 37°C (the asterisk denotes the labeled strand). The homoduplex 60-mer substrate (G•C*) was obtained by annealing the 60-mer G with the labeled 60-mer C-F oligonucleotide. Where recombinant proteins were used, the standard AP endonuclease reaction buffer contained 50mM Tris pH 8.0, 1mM DTT, 1mM EDTA and 25 ng/µl BSA. Whenever the *P. aerophilum* WCE was used, the same buffer described above was employed and reactions were terminated by the addition of 10 µl of stop solution and incubated for 30 min at 37°C. Incubation times and temperatures varied as indicated in the figures. After precipitation, the DNA pellets were dissolved in 90% formamide supplemented with 50mM $NaBH_4$ in order to prevent spontaneous hydrolytic cleavage of the labile AP-sites prior to analyzing the samples as described (10).

DNA polymerase assays- The primer-extension ability of the DNA polymerases was compared using a substrate prepared by annealing the 5'-labeled 23-mer-F primer with the 60-mer G template oligonucleotide (listed in Table I), illustrated in Figure 3A. Unless otherwise mentioned, the standard reaction mixtures contained, in a 20 µl volume, 50mM Tris pH 8.5, 2mM $MgCl_2$, 25 ng/µl BSA, 0.2mM EDTA, 1mM DTT, 100µM dNTP, 1 pmol of the

labeled substrate and purified recombinant DNA polymerases. Incubation conditions varied as indicated in the figures. The reaction products were separated on denaturing 15% polyacrylamide gels and the bands were visualized on a FluoroImager (Storm 860, Molecular Dynamics) as described previously (10). For 3'→5 exonuclease activity on single strands (ss), the degradation of the labeled 23-mer-F primer was analyzed. 3→5 exonuclease activity on double strands (ds) was measured using the same substrate as for DNA polymerizing activity. Intrinsic 3→5 proofreading features of different DNA polymerases were assayed in the absence of dNTPs. A 20mM stock solution of the polymerase inhibitor aphidicolin was prepared by dissolving 1 mg of aphidicolin (Calbiochem) in 150 µl 100 % DMSO. In DNA polymerase inhibition reactions, 2mM aphidicolin was pre-incubated with the DNA polymerase prior to adding the enzyme to the reaction mixtures. A 5mM stock solution of ddCTP was purchased from Amersham and used in a 10:1 ratio with 100µM dNTPs in standard reaction mixtures. Wherever indicated, 500 ng of recombinant purified *Pa*-PCNA1 was included in the reactions, corresponding to approximately 5 pmol of the functional trimeric complex (23).

Electrophoretic Mobility Shift Assay (EMSA)- In EMSA reactions (10 µl), *Pa*-UDGb (5 pmol), *Pa*-PolB3 (2 pmol), and *Pa*-PolB2 (1 and 2 pmols) were incubated in the presence or absence of *Pa*-PCNA1 (500 ng) with 1 pmol of labeled 60-mer substrates (1nt-gap or G•C*) and 10 pmol of unlabeled G•C competitor DNA in 25mM Tris pH 8.0, 1mM DTT, 0.5mM EDTA and 5% glycerol for 30 min at 4°C. The protein-DNA complexes were separated by electrophoresis on 5% native polyacrylamide gels in 0.5 x TBE at 4°C and analyzed as described previously (11).

DNA Ligase Adenylation Assay- 5 pmol of recombinant *Pa*-DNA-Ligase was

incubated with 18.5 kBq of [α-^{32}P]ATP (Amersham) in a buffer containing 60mM Tris pH 8.0, 10mM $MgCl_2$, 5mM DTT and 50 ng/µl BSA for 15 min at different temperatures as indicated in Figure 4A. The reactions (10 µl) were stopped by boiling in 1 x SDS loading buffer and analyzed by electrophoresis by 10% SDS-PAGE. After the gel was dried, adenylated proteins were detected by autoradiography. T4 DNA Ligase was used as a positive control.

DNA Ligase Assay- The radioactively-labeled substrate, illustrated in Figure 4B, was prepared as follows: 1 µg of a 24-mer oligodeoxythymidilate (dT) was phosphorylated at its 5'-terminus by incubation with 1.85 MBq of [γ-^{32}P]ATP (Amersham), 50 pmol of ATP and 30 units of T4 Polynucleotide Kinase (New England Biolabs) for 45 min at 37°C. The kinase was heat-inactivated and the products purified by centrifugation through a G-25 Sephadex Spin Column (Roche). The radioactively-labeled oligo-dT was then annealed with polydeoxadenylic acid (poly-dA; approximately 270 nucleotides in length; Amersham) in the presence of an excess of cold 5'-phosphorylated oligo-dT.

Ligation reaction mixtures (10 µl) containing 1 x T4 DNA Ligase buffer, 500 fmol of the labeled substrate poly-dA:oligo-dT and purified recombinant *Pa*-DNA-Ligase where incubated for 15 min at 55°C. T4 DNA Ligase (40 units) was used as the positive control, at 37°C. The reactions were terminated by adding 30 µl of 90 % formamide dye loading buffer. The samples were then heated for 5 min at 95 °C and electrophoresed through a 10 % denaturing gel. The gel was dried and the reaction products were detected by autoradiography.

Fluorescently-labeled substrates containing a ligatable nick were used to assay nick-joining activity of the recombinant *Pa*-DNA-Ligase and of the DNA ligase(s) present in *P.*

aerophilum WCE. The labeled 24-mer C-F (or 24-mer T-F) and the unlabeled 36-mer oligos were annealed with the complementary 60-mer G oligo (see Table I). For an illustration of the substrates see figure 4C. The standard reaction mixtures (20 µl) contained 50mM Tris pH 8.0, 25 ng/µl BSA, 0.2mM EDTA, 1mM DTT, 1 pmol of the nicked substrates and either 5 pmol of recombinant *Pa*-DNA-Ligase or 20 µg of *P. aerophilum* WCE. Co-factors (5mM $MgCl_2$ and 2mM ATP) and 500 ng of recombinant *Pa*-PCNA1 were added as indicated in the figures and the reactions were incubated for 1 h at 60°C. T4 DNA Ligase was used as a positive control, and was incubated with the substrates in the supplied T4 DNA Ligase buffer for 1 h at 37°C.

Partial and complete BER reconstitution assays using 1nt-gap, AP-site and G•U mismatch-containing substrates and P. aerophilum recombinant proteins- The 1-nt gap substrate (see Figure 1A) to assay the combined action of the *P. aerophilum* DNA polymerases (*Pa*-PolB2 and *Pa*-PolB3) and the *Pa*-DNA-Ligase was prepared as described above. All reaction mixtures contained, in a 20 µl volume, 50mM Tris pH 8.0, 10mM $MgCl_2$, 25 ng/µl BSA, 0.2mM EDTA, 5mM DTT and 2mM ATP. 1 pmol of the substrate was incubated with *Pa*-PolB2 (0.5 pmol) or *Pa*-PolB3 (0.1 pmol) in the presence or absence of *Pa*-DNA-Ligase (5 pmol) and *Pa*-PCNA1 (500 ng) for 30 min at 60°C. The reaction was initiated by the addition of 20µM dCTP or 20µM of each dNTPs, respectively.

The repair of AP-sites by purified recombinant *P. aerophilum* proteins was monitored by using substrates containing either a normal or a reduced AP-site as illustrated in Figure 5B. For normal AP-sites, the G•AP* 60-mer substrate was prepared as described above. The substrate containing a reduced AP-site (RAP) was prepared as follows: 40 pmol of the G•U* 60-mer were first incubated in a total volume of 20 µl with 4 units of *E. coli* UDG for 30

min at 37°C to produce AP-sites, which were then reduced for 15 min on ice by the addition of 0.1M $NaBH_4$. The resulting sample (H 1μM of the G•RAP* 60-mer) was then filtered through a G-25 Sephadex Spin Column (Roche) for 5 min at 4°C at 2000 rpm to wash away the reducing agent. Both substrates (1 pmol) were then incubated for 30 min at 60°C in a reaction mixture (20 μl) containing 50mM Tris pH 8.0, 25 ng/μl BSA, 7.5mM $MgCl_2$, 0.5mM EDTA, 2mM DTT and 2mM ATP with recombinant *P. aerophilum* proteins as indicated in figure 5B. Reactions were initiated by the addition of 20μM dCTP or 20μM of each dCTP, dGTP and dATP (d3TPs). dTTP was omitted from the reaction to allow a maximum incorporation of 5 nucleotides, according to the oligonucleotide sequence of the 60-mer G-template as denoted in Table I, thereby limiting strand displacement DNA synthesis by *Pa*-PolB2.

The substrate used for the complete *in vitro* reconstitution of uracil-BER by recombinant *P. aerophilum* proteins was prepared by annealing the labeled 60-mer U-F oligonucleotide with its complementary strand (50-mer G), to generate a heteroduplex containing a single G•U mismatch. As illustrated in figure 5C, the unlabeled strand lacks 10 nucleotides at the 5'-end to clearly distinguish a fully-repaired 60-mer product form a 50-mer product resulting from a strand displacement synthesis by a DNA polymerase. The reactions (20 μl) were carried out in two steps, in a buffer containing 50mM Tris pH 8.0, 5mM $MgCl_2$, 2mM DTT, 0.2mM EDTA, 25 ng/μl BSA and 20μM of each dNTPs. First, 1 pmol of the substrate was incubated with *Pa*-UDGb (1 pmol) and *Pa*-EndoIV (0.2 pmol) for 15 min at 60°C to remove the aberrant base (uracil) and cleave the resulting AP-site. In a second step, 2mM ATP and pre-mixes of different combinations of *Pa*-PCNA1 (500 ng), *Pa*-DNA-Ligase (5 pmol) and either *Pa*-PolB2 or *Pa*-PolB3 (0.5 pmol, respectively) were added and the reaction mixtures

were further incubated for 30 min at 60°C. The reactions were terminated by adding the precipitation mixture (5 µg tRNA, 300mM sodium acetate pH 5.2 and 3 volumes of ice-cold ethanol) and putting the samples immediately at –20°C. The samples were analyzed as described previously (10).

RESULTS

Base Excision Repair activities supported by P. aerophilum extracts- In our earlier studies, we demonstrated that whole cell-free extracts (WCE) of *P. aerophilum* could catalyze the removal of uracil from a G•U mismatch and the subsequent cleavage of the resulting AP-site (10). As this results in the generation of a single nucleotide gap through the removal of the abasic sugar-phosphate, we wanted to test the capacity of *P. aerophilum* WCE to repair this latter lesion (Figure 1A). Upon addition of deoxyribonucleotide triphosphates and magnesium, we observed preferential incorporation of a single nucleotide into the one-nucleotide gap in the oligonucleotide substrate (lane 3). A faint product band of 60 nucleotides was also observed (lane 3), which could have arisen either through a complete displacement of the downstream 36-mer oligonucleotide by a DNA polymerase, or through nick-joining mediated by a pre-adenylated DNA ligase in the extract. Surprisingly, this putative ligation reaction was not further enhanced in the presence of ATP (lane 4). Addition of T4 DNA Ligase to the reaction gave rise to greater amounts of the 60-mer product, but this enzyme was also inhibited by ATP (cf. lanes 5 and 6).

We next wanted to see whether the extracts could repair a G•U mispair in an oligonucleotide substrate. This heteroduplex was constructed such that it allowed us to distinguish between strand displacement and DNA repair synthesis. Thus, strand displacement resulted in a 50-mer product, because the lower strand, serving as a template for the DNA polymerase(s), was recessed at the 5'-end. In contrast, BER gave rise to a 60-mer product (see scheme at the top of Fig. 1A). When this substrate was incubated with the extracts, we could observe either the formation of a 23-mer product (lane 8), which arose through the removal of the aberrant base in the absence of added dNTPs and Mg^{2+}, or of a

24-mer (lane 9), which represented the 23-mer extended by a polymerase activity present in the extracts, provided that the extracts were supplemented with the necessary co-factors. The repaired 60-mer product was also evident (lane 10), and increased in abundance in the presence of T4 DNA Ligase (lanes 11 and 12). As the WCE of *P. aerophilum* was apparently able to carry out all the steps of BER, we set out to characterize the enzymes involved.

Identification of BER genes in P. aerophilum- We showed earlier that *P. aerophilum* possessed at least three different DNA glycosylases able to remove uracil from a G•U mismatch *in vitro* (9-11). The action of a monofunctional uracil-DNA glycosylase (e.g. *Pa*-UDGa or *Pa*-UDGb) generates an AP-site, which is incised by an AP endonuclease such as APE1 in man, or APN1 in *S. cerevisiae*. Both enzymes are the major constitutively-expressed AP endonucleases in these organisms and are homologous to the *E. coli* exonuclease (Exo) III and endonuclease (Endo) IV, respectively, the founding members of the two AP endonuclease families (28). Despite their similar enzymatic activities, ExoIII and EndoIV share no primary sequence identity and are structurally unrelated. By searching the genome database of *P. aerophilum* for open reading frames (ORFs) encoding putative homologues of both types of AP endonucleases, we failed to identify an ExoIII homologue, but ORF PAE3257 was found to encode a polypeptide with significant homology to *E. coli* EndoIV (*Ec*-EndoIV). Endonuclease IV proteins use conserved histidine (H), aspartate (D) and glutamate (E) side chains to form a trinuclear zinc cluster (29). As the 275-amino acid ORF identified contained all the latter metal-binding residues, we named its protein product *Pa*-EndoIV (Fig. 1B).

The removal of uracil from a G•U mismatch and the cleavage of the resulting AP-site on its 5'-side gives rise to a single-strand break where the upstream fragment is terminated

with a 3'-hydroxyl group and the downstream fragment has a deoxyribose-phosphate (dRP) group at its 5'-terminus. In mammalian short-patch BER, DNA polymerase-β (pol-β) removes the dRP by its associated AP-lyase activity and simultaneously extends the 3'-terminus by one nucleotide (17, 30). The dRP-lyase step, which is rate-limiting in BER (31), was shown to be carried out by the 8 kDa N-teminal domain of pol-β, a member of the X-family of DNA polymerases. The sequence of pol-β was used to search for putative homologues in the *P. aerophilum* sequence database, albeit without success. Indeed, only family-B DNA polymerases, thought to be involved primarily in DNA replication, have been identified in crenarchaeal organisms to date (32). As shown in Figure 1C, the *P. aerophilum* genome encodes three family-B DNA polymerases, one each of the B1, B2, and B3 subfamilies, denoted as *Pa*-PolB1, *Pa*-PolB2, and *Pa*-PolB3, respectively (5, 33). Interestingly, while *Pa*-PolB1 and *Pa*-PolB3 contain all known signatures of replicative DNA polymerases, the *Pa*-PolB2 sequence is substantially shorter and lacks several conserved motifs, one of which encodes a putative proofreading exonuclease (Fig. 1C). However, the 553 amino acid ORF contained the two invariant aspartates that form the catalytic dyad of all known family-B DNA polymerases. These sequence differences suggested that *Pa*-PolB2 might be involved in DNA repair, while *Pa*-PolB1 and *Pa*-PolB3 might be the replicative polymerases (34).

In the final step of BER, the nick remaining after the polymerase had filled the gap needs to be sealed. In human short-patch BER, this role is accomplished by DNA ligase III, in an ATP-dependent manner. Homology searches of the *P. aerophilum* genome using the catalytic domain of human DNA ligase III revealed the existence of a single ORF (PAE833) encoding an ATP-dependent DNA ligase. Thus, like all known archaeabacterial ligases, the

P. aerophilum enzyme is also ATP- rather than NAD-dependent (see (35, 36) for reviews). As shown in Figure 1D (upper panel), the domain organization of *Pa*-DNA-Ligase is quite similar to human DNA ligase I, and contains all six conserved sequence motifs typical of ATP-dependent DNA ligases (lower panel). However, *Pa*-DNA-Ligase lacks an N-terminal extension that contains important functional motifs in mammalian DNA ligases (37).

The *P. aerophilum* proteins *Pa*-EndoIV, *Pa*-PolB2 and *Pa*-DNA-Ligase were overexpressed in *E. coli*. In addition, we expressed also *Pa*-PolB3 and *Pa*-PCNA1. As shown in Figure 1E, the proteins could be purified to apparent homogeneity, with the exception of *Pa*-DNA-Ligase. This protein migrated through the denaturing gel with a molecular mass of 64 kDa, close to the calculated value of 67 kDa. However, the fraction contained also two other, minor species with apparent molecular masses of ~160 and 50 kDa. As all three bands could be visualized on Western blots using an anti-His_6 antibody (data not shown), all three species originated from the transfected vector. It is possible that the smaller protein is a truncated version of *Pa*-DNA-Ligase, which could have arisen through heat treatment during the workup. The nature of the larger species is unclear. It could be due to incomplete denaturing of the protein, or represent a post-translationally modified form of *Pa*-DNA-Ligase. *Pa*-PCNA1 migrated through polyacrylamide gels with an apparent molecular mass of ~37 kDa, rather than the calculated 28 kDa, as noted previously (23).

Analysis of Pa-EndoIV enzymatic activities- The cleared lysate of *E. coli* BL21(DE3) cells transfected with the *pET28-PaEndoIV* plasmid construct was adsorbed on a Ni-NTA column and the retained proteins were eluted with 250mM imidazole (Experimental Procedures). Both the soluble fraction and the Ni-NTA eluate catalyzed the cleavage of a 60-mer oligonucleotide duplex containing a single AP-site (Fig. 2A lanes 5

and 8). The activity was stable up to 70°C (lanes 6 and 9), however, as the *E. coli* AP endonuclease IV is also stable at this temperature (38), we had to formally eliminate the possibility that the activity we observed was due to a co-purifying bacterial protein. To this end, we tested the corresponding fractions obtained from transformation of the host BL21 cells with the empty pET vector. While the AP endonuclease activity was indeed present in the soluble fraction (lane 3) even after heat treatment (lane 4), no activity was detected after Ni-NTA purification (lane 7). The activity present in the Ni-NTA column fractions (lanes 8 and 9) was therefore due to the overexpressed *Pa*-EndoIV.

In order to characterize the *Pa*-EndoIV more closely, we purified it further using Hi-Trap Q and Hi-Trap Heparin FPLC columns. The enzymatic activity of this protein could be shown to increase with increasing temperature (Figure 2B), as expected of an enzyme encoded by an organism with an optimal growth temperature of 100°C. The experiment also demonstrated that the Zn^{2+}-dependent *Pa*-EndoIV endonuclease resists inactivation by 5mM EDTA, unlike the Mg^{2+}-dependent AP endonucleases belonging to the ExoIII family (39).

Pa-UDGa and *Pa*-UDGb are both monofunctional DNA glycosylases that generate AP-sites in the 60-mer oligonucleotide duplex G•U, but are unable to convert these into strand breaks, because they lack an associated DNA lyase activity (Fig. 2C, lanes 2 and 3). The 60-mer substrate could be cleaved at these sites by β-δ-elimination with hot alkali (Fig. 2C, lanes 4 and 5), which gave rise to a labeled 23-mer fragment terminated with a 3'-phosphate group. *Pa*-EndoIV also cleaved the AP-sites created by both these enzymes. However, because the 60-mer substrate was cleaved on the 5' side of the abasic residue by hydrolysis of the phosphodiester linkage, the labeled fragment was terminated with a 3'-

hydroxyl group (lanes 6 and 7), and therefore migrated slower through the polyacrylamide gel. The lower efficiency of AP-site cleavage seen in lane 7 (as compared to lane 6) might be explained by the fact that *Pa*-UDGb, unlike *Pa*-UDGa, strongly binds to AP-sites (11), thereby shielding these sites from cleavage by *Pa*-EndoIV.

Incubation of the labeled G•U 60-mer substrate with *P. aerophilum* WCE (Figure 2D, lanes 3-5) revealed that the cleaved product migrated at the same position as that generated by the combined action of purified recombinant *Pa*-UDGa and *Pa*-EndoIV (Figure 2D, lane 2). This processing was mediated by proteins in the extract that did not require Mg^{2+} (lane 3) and that exhibited substantial thermostability (lane 4), which is again consistent with a role of an EndoIV-type AP endonuclease in the processing of AP sites *in vivo*. Interestingly, when the extracts were supplemented with dNTPs and Mg^{2+}, the DNA polymerase(s) present in the extract extended the cleaved 23-mer by a single nucleotide; very little strand displacement could be detected under our experimental conditions (Figure 2D, lane 6). However, the labeled extended 24-mer product failed to be ligated to the downstream 36-mer by the endogenous DNA ligase (Fig. 2D, lane 6; see also Fig. 1A).

Pa-PolB2 DNA polymerase is a functional homologue of pol-β - To study the enzymatic activities of the purified recombinant *P. aerophilum* DNA polymerases from *Pa*-PolB2 and *Pa*-PolB3, we used a fluorescently-labeled primer-extension substrate shown in Fig. 3A. In the first assay, we wanted to test whether the recombinant *Pa*-PolB2 indeed possesses DNA polymerase activity, as its amino acid sequence contains several amino acid substitutions in the highly-conserved polymerase motifs (Fig. 1C). As shown in Fig. 3A, the fraction of soluble extracts from *E. coli* BL21 cells overexpressing *Pa*-PolB2 that was retained on the Ni-NTA column contained DNA polymerase activity, which was stimulated

at higher temperatures (lanes 6-8). The corresponding fraction from cells transformed with the empty vector was devoid of this activity (Fig. 3A, lane 5) and we therefore concluded that the observed polymerase activity was intrinsic to the overexpressed *Pa*-PolB2.

In the second assay, we set out to test for the presence of a 3'→5 proofreading exonuclease activity. The purified recombinant *Pa*-PolB2 and *Pa*-PolB3, together with *Taq* (family A) and *Pfu* (family B) polymerases, were incubated with the labeled single-stranded 23-mer primer, or with the labeled primer-extension substrate (Fig. 3A) in the presence or absence of dNTPs. As shown in Fig. 3B, *Taq* pol and *Pa*-PolB2 were devoid of a 3→5 exonuclease activity on both substrates in the absence of dNTPs (lanes 1,2 and 9,10, respectively). In contrast, *Pfu* pol and *Pa*-PolB3 efficiently degraded the labeled primer in both single-stranded and double-stranded substrates (lanes 4,5 and 14,15, respectively). This was anticipated, as *Pa*-PolB3 shares 78% amino acid identity with a recently-characterized *Pyrobaculum islandicum* DNA polymerase, which was reported to contain a 3'→5 exonuclease activity (40). As anticipated, in the presence of dNTPs, all four polymerases switched to the primer extension mode (lanes 3,6,11,16). In the presence of 2mM aphidicolin, a generic inhibitor of eukaryotic replicative polymerases and several polymerases of the archaeal family B (40), only *Pfu* pol and *Pa*-PolB3 were inhibited (lanes 8 and 18, respectively). *Pa*-PolB2 (lane 13) was insensitive to this drug.

In addition to their sensitivity to aphidicolin, the three major mammalian polymerases involved in DNA replication (α, δ and ε) do not incorporate dideoxyribonucleotide monophosphates (ddNMPs) into DNA. *Pfu* pol, a representative member of the archaeal B-family of DNA polymerases, behaved similarly (Fig. 3C, lanes 8,9). In contrast, the human pol-β is not inhibited by aphidicolin (Fig. 3C, lane 4), but incorporates ddNMPs into DNA

quite efficiently (20, 41); the primer extension reaction is thus inhibited by these substances (Fig. 3C, lane5). In all these assays, *Pa*-PolB2 resembled pol-β. It remained unaffected by 2mM aphidicolin (lane 12), but was inhibited by dideoxynucleoside triphosphates at a ddCTP/dCTP ratio of 10, albeit not as efficiently as pol-β (cf. Lanes 5 and 13).

Next, we wanted to test whether DNA synthesis by *Pa*-PolB2 is affected by the recombinant *P. aerophilum* polymerase processivity factor *Pa*-PCNA1. As shown in Fig. 3D, *Pa*-PCNA1 appeared to slightly inhibit *Pa*-PolB2 (lanes 2-5), while DNA synthesis catalyzed by *Pa*-PolB3 was greatly stimulated by the presence of the sliding clamp (lanes 6-9). This finding was consistent with the fact that no putative PCNA-binding motif could be identified in the amino acid sequence of *Pa*-PolB2. In contrast, *Pa*-PolB3 contains this motif (Fig. 1C) and could be shown to interact with *Pa*-PCNA1 *in vitro* (23).

Finally, we wanted to test whether these archaeal polymerases displayed binding affinity towards a double-stranded 60-mer oligo containing a single nucleotide gap. This substrate, which is generated during the initial steps of BER by the cleavage of an AP-site and the subsequent removal of the baseless sugar-phosphate (dRP), was bound by *Pa*-UDGb (Fig. 3E, lane 6). *Pa*-UDGb interacted weakly also with the G•C homoduplex DNA (Fig. 3E, lane 2) and more strongly with AP-sites (10). *Pa*-PolB2 also bound to the gapped-substrate in an EMSA assay (Fig. 3E lanes 8 and 9), albeit with an affinity lower than *Pa*-UDGb (lane 6). Interestingly, the mobility of the retarded band in the polyacrylamide matrix was slightly lower in the presence of *Pa*-PCNA1 (lanes 10 and 11), suggesting that the DNA, *Pa*-PolB2 and *Pa*-PCNA1 formed a ternary complex. This finding was rather unexpected, in view of the fact that *Pa*-PolB2 lacks a PCNA binding motif and that its polymerase activity was not stimulated by the sliding clamp, in contrast to *Pa*-PolB3, which interacts and is stimulated by

Pa-PCNA, yet could be seen to form no stable complex on this DNA substrate (lanes 12, 13). However, in the light of data presented below, *Pa*-PCNA1 may have a role in BER that is distinct from its role as a processivity factor in DNA synthesis.

Catalytic Properties of Pa-DNA-Ligase- The first step of the ligation reaction involves a nucleophilic attack of the lysine residue within the active site motif I (KYDGER in *Pa*-DNA-Ligase, Fig. 1D, lower panel) on the α-phosphate of ATP, which results in the formation of a lysine-AMP adduct. As shown in Figure 4A, a radiolabeled DNAligase-adenylate adduct was formed in the presence of ATP, which co-migrated with the *Pa*-DNA-Ligase protein in SDS-PAGE. As anticipated, the adenylation of *Pa*-DNA-Ligase was more efficient at 70°C than at 37°C. In the subsequent steps of the reaction, the adenylate moiety is transferred from the ligase to the 5'-phosphate group of the nicked DNA to give rise to a highly-reactive pyrophosphate intermediate. The two DNA fragments are joined by a nucleophilic attack on the pyrophosphate by the 3'-hydroxyl moiety of the upstream DNA fragment. The efficiency of *Pa*-DNA-Ligase in joining single-strand breaks in double-stranded DNA was compared with that of T4 DNA Ligase in an *in vitro* assay using a radioactively-labeled substrate. As shown in Fig. 4B, the latter protein was substantially more efficient than *Pa*-DNA-Ligase. However, it should be remembered that the incubation was carried out at sub-optimal temperature (55°C) for the archaeal enzyme, due to the relatively low melting temperature of the substrate.

To compare the nick-joining activities of the recombinant *Pa*-DNA-Ligase with those of the enzyme contained in *P. aerophilum* WCE, we designed two fluorescently-labeled substrates containing either a 3'-matched (G•C) or 3'-mismatched (G•T) base pair at the nick. While T4 DNA Ligase, known to tolerate mismatches on either side of the nick

(42), was able to seal both substrates, only the matched substrate was ligated by the recombinant *Pa*-DNA-Ligase and by the native enzyme in the cell extracts (Figure 4C). The matched substrate was quite efficiently ligated even when ATP was omitted from the reaction mixture (lane 3), most likely due to the presence of pre-adenylated recombinant *Pa*-DNA-Ligase in the purified fraction. This phenomenon was observed also when the WCE was used (see Fig. 1A, lane 2; Fig. 4C, lane 7). Unexpectedly, *Pa*-DNA-Ligase was able to seal the mismatched substrate, at least to a small extent, in the presence of *Pa*-PCNA1 (Fig. 4C, lane 13). This effect is apparently not due to a PCNA-mediated stimulation of ligase activity, as no such effect was observed when the matched substrate was used (cf. lanes 4 and 5). We postulate that *Pa*-PCNA1 may assist ligation by binding to the substrate in the vicinity of the termini, which may keep the ends from "fraying" due to the presence of the mispair.

Reconstitution of base excision repair using purified recombinant P. aerophilum proteins- As all three DNA polymerases of *P. aerophilum* belong to family B, and as we failed to identify an ORF encoding a pol-β homologue, i.e. an X-type DNA polymerase, we first wanted to see which *P. aerophilum* polymerase catalyzes the repair of a single-strand break most efficiently. Using a substrate containing a one nucleotide gap opposite a G residue (see Fig. 1A), both *Pa*-PolB2 or *Pa*-PolB3 extended the 23-mer primer by one nucleotide, providing that the reaction was supplemented with dCTP only (Figure 5A, lanes 2 and 3, respectively). In the presence of *Pa*-DNA-Ligase, the repaired 60-mer oligo was produced in a slightly higher yield in the reaction containing *Pa*-PolB2 (cf. lanes 4 and 5), and this trend was even more evident in the presence of *Pa*-PCNA1 (cf. lanes 6 and 7), when the intensity of the bands due to the unligated 24-mers is considered. When the experiment was carried out in the presence of all four dNTPs, *Pa*-PolB2 was more efficient than *Pa*-PolB3 in

catalyzing limited strand displacement (cf. lanes 8 and 9). However, in the presence of *Pa*-DNA-Ligase, the band corresponding to the repaired 60-mer, which arose most likely through the extension of the 23-mer primer by a single nucleotide and subsequent ligation, was more prominent in the reaction containing *Pa*-PolB2 than that containing *Pa*-PolB3 (cf. lanes 10 and 11). In the presence of *Pa*-PCNA1, the strand displacement activity of *Pa*-PolB3 increased, while that of *Pa*-PolB2 appeared to have been inhibited to some extent (cf. lanes 12 and 13). In the presence of *Pa*-PCNA1 and *Pa*-DNA-Ligase, both polymerases yielded similar amounts of the 60-mer product (lanes 14 and 15).

The results of the experiments carried out up to this point do not rule out the participation of either polymerase in BER. We decided to examine *Pa*-PolB2 in greater detail, as its enzymatic properties resemble more those of pol-β, which is a principal player in mammalian BER. In particular, we wanted to see how this enzyme addresses its real BER substrate, namely, a double-stranded DNA molecule containing a cleaved AP-site. In the processing of such a substrate, the enzyme has not only to extend the 3'-terminus by a single nucleotide; it must concomitantly catalyze the removal of a dRP moiety that is attached to the 5'-terminus of the downstream DNA fragment to generate a 5'-phosphate that can be subsequently ligated to the 3'-hydroxyl group of the extended upstream fragment. To test our enzymes on this type of substrate, we generated oligonucleotide duplexes containing either a normal (AP) or a reduced (RAP) abasic sites (see diagram in Fig. 5B and Experimental Procedures for details). After endonucleolytic cleavage, the RAP substrate is resistant to β-elimination, and its "repair" can be accomplished only by displacement of the entire downstream 36-mer (see above), or by excision mediated by a 5→3 exonuclease (FEN1), followed by gap-filling (pol-δ) and ligation (long-patch repair). In contrast, the unreduced

dRP moiety can be removed also by a β-elimination reaction, as described for the mammalian system (30). As shown in Figure 5B, the AP-site was sensitive to this reaction even in the absence of added enzymes, especially at higher temperatures (upper panel, lanes 1 and 2), and this reaction was accelerated in the presence of *Pa*-PolB2 and *Pa*-DNA-Ligase [lane 3, see also ref. (43)]. As expected, the RAP-site was stable under these experimental conditions (lower panel, lanes 1-3). In the presence of *Pa*-EndoIV (lanes 4), both sites were efficiently cleaved. Note that the product band generated by *Pa*-EndoIV migrated slightly faster through the polyacrylamide gel than the product resulting from the spontaneous β-elimination reaction. This is because the former product has a 3'-OH terminus, whereas the latter has an α,β-unsaturated aldehyde at its 3'-terminus (41).

In the presence of *Pa*-PolB2 and dCTP, the 3'-termini of both substrates were extended by a single nucleotide (lanes 5), but only the AP substrate was repaired to the 60-mer product (lane 6) in the presence of *Pa*-DNA-Ligase. This shows that the abasic sugar-phosphate (dRP residue) was removed to produce ligatable ends. No ligation took place on the RAP substrate, which shows that the 5-terminus of the downstream 36-mer was blocked by the reduced dRP moiety. Moreover, this result shows that *Pa*-PolB2 possesses no 5→3 exonuclease activity. In the presence of dCTP, dATP and dGTP (d3TP), *Pa*-PolB2 catalyzed a limited strand displacement reaction, extending the 23-mer primer by several nucleotide residues (according to the sequence of the oligonucleotide shown in Table I, a maximum extension by 5 residues was possible under these conditions). However, the 24-mer was the most prominent (upper panel, lane 7), and was the only extension product that was converted to the 60-mer (upper panel, lane 8). The strand displacement products obtained with both AP and RAP substrates were not converted to the 60-mers, which shows that *Pa*-PolB2

possessed no intrinsic flap endonuclease activity that could remove the 5'-overhangs.

The results presented above implied that *Pa*-PolB2 might be the functional *P. aerophilum* homologue of the mammalian pol-β. It appeared to possess a dRP-lyase activity similar to that encoded in the N-terminal 8 kDa domain of the human DNA polymerase β (44), even though we failed to identify a helix-hairpin-helix motif in *Pa*-PolB2 that might be indicative of such an activity. However, the enzyme contains 10 lysine residues within the 150 N-terminal amino acids, one or more of which could act as nucleophile(s) during the β-elimination reaction. Moreover, the EMSA experiment shown in Figure 3E indicated that *Pa*-PolB2 can form stable complexes with gapped DNA substrates. A similar property was demonstrated for the 8-kDa domain of pol-β (44).

However, when this experiment was carried out with *Pa*-PolB3, an identical result was obtained (data not shown), implying that the dRP moiety could be removed under our reaction conditions in the presence of either polymerase. We therefore tested both polymerases in the subsequent experiment, in which we wanted to test whether the recombinant *P. aerophilum* BER proteins described above could catalyze the repair of a G•U mismatch in the 50-mer/60-mer oligonucleotide substrate used above (Fig. 1A). As shown in Fig. 5C (lane 1), the process was absolutely dependent on *Pa*-UDGb, which catalyzed removal of the uracil residue and generated thus an AP-site that could be cleaved by *Pa*-EndoIV to give rise to a 23-mer product (lane 3). We chose *Pa*-UDGb rather than one of the other uracil DNA-glycosylases for the initiation of the repair process, because we believe that *Pa*-UDGb is most likely to play this role also *in vivo* (see Discussion). In the absence of *Pa*-DNA-Ligase and *Pa*-PCNA1, both polymerases, *Pa*-PolB2 or *Pa*-PolB3, catalyzed substantial strand displacement synthesis, but in the presence of *Pa*-PCNA1, *Pa*-PolB2 gave

rise primarily to the 24-mer gap-filled product (lane 6), whereas *Pa*-PolB3 yielded predominantly the 50-mer product by strand displacement (lane 7). In the presence of *Pa*-DNA-Ligase and absence of *Pa*-PCNA1, both polymerases yielded similar amounts of the 60-mer product (lanes 8 and 9). This situation changed when *Pa*-PCNA1 was also present. In the assays containing *Pa*-UDGb, *Pa*-EndoIV, *Pa*-PCNA1, *Pa*-DNA-Ligase and *Pa*-PolB2, the main product was the 60-mer resulting from BER (lane 10). When *Pa*-PolB2 was substituted with *Pa*-PolB3, the ratio of the BER and strand displacement products was approximately equal (lane 11). Although these results do not exclude either polymerase from participating in BER in *P. aerophilum*, the similarity of the biochemical properties of *Pa*-PolB2 and pol-β, as well as its lack of stimulation and interaction with *Pa*-PCNA1, suggest that *Pa*-PolB2 is more likely to be involved in DNA repair, rather than in DNA replication.

DISCUSSION

In the experiments described above we showed that the genome of *P. aerophilum* (5) encodes a full set of BER proteins. In addition to several DNA glycosylases (9-11), ORFs encoding one AP endonuclease, three DNA polymerases of the B family and one DNA ligase could be identified by sequence homology searches. In addition, the genome encodes also two homologues of the polymerase processivity factor PCNA (23). We expressed *Pa*-EndoIV, *Pa*-PolB2, *Pa*-PolB3 and *Pa*-DNA-Ligase and showed that the purified recombinant proteins, expressed in *E. coli*, could repair a G•U mispair in an oligonucleotide substrate to a G•C pair, providing that the reconstituted reaction contained also a uracil-DNA glycosylase such as *Pa*-UDGb. As anticipated, all the individual components of this pathway were considerably thermostable, and the repair reaction itself proceeded with greater efficiency at 60ÚC than at lower temperatures. Thus, *P. aerophilum* possesses all the enzymatic functions necessary for efficient base excision repair, a DNA metabolic pathway of substantial importance, especially in organisms growing at high temperatures.

Taking the repair of uracil arising through spontaneous hydrolytic deamination of cytosine as an example, earlier work showed that the *in vitro* BER process could in this case be initiated by any one of at least three uracil-DNA glycosylases: *Pa*-UDGa (10), *Pa*-UDGb (11) or *Pa*-MIG (9). However, we would argue that the enzyme most likely to initiate this repair process *in vivo* would be *Pa*-UDGb. As *Pa*-UDGa could be shown to interact with *Pa*-PCNA1, this glycosylase might be more likely to act in the repair of uracil residues incorporated into the newly-synthesized strand in the form of dUMP during DNA replication (45). *Pa*-MIG is a DNA glycosylase that can address both G•U and G•T mispairs, the latter thought to arise through the deamination of 5-methylcytosine residues in DNA (9). Its

participation in short patch BER of G•U mispairs cannot be ruled out, but its low activity and abundance in cell-free extracts of *P. aerophilum* (10) might imply that this enzyme plays only a secondary role to the more active and more abundant *Pa*-UDGb *in vivo*. It was for this reason that we selected the latter glycosylase as the initiating enzyme in our BER reconstitution experiments.

The AP-site arising through the removal of uracil by *Pa*-UDGb has to be incised at its 5'-end by an AP endonuclease. *Pa*-EndoIV appears to be the only enzyme of this kind encoded in the genome of *P. aerophilum*, and it is thus likely that this protein indeed participates in BER in this organism. *Pa*-EndoIV is required also for the processing of AP-sites generated by the action of glycosylases/lyases, such as enzymes of the Nth family (46) that cleave the DNA backbone by β-elimination concomitantly with base removal. The α,β-unsaturated aldehyde generated by these enzymes blocks the 3-terminus of the upstream fragment such that the repair polymerase cannot initiate the gap-filling reaction until this moiety is cleaved off. *Pa*-EndoIV can indeed catalyze the removal of these residues *in vitro* (data not shown).

Following the action of *Pa*-UDGb and *Pa*-EndoIV, the 3'-terminus of the incised strand has a free hydroxyl group, and can thus serve as a primer for the repair polymerase in a gap-filling reaction. However, the end of the extended primer cannot be ligated to the 5'-terminus of the incised strand until the baseless sugar-phosphate (dRP) that is blocking this site is removed. This reaction can take place spontaneously to some extent, especially at elevated temperatures (Fig. 2C), but it is unlikely that any organism would rely on this process. In mammals, the β-elimination reaction that removes the dRP residue is catalyzed by the N-terminus of pol-β (47). We did not directly test the dRP lyase activity of *Pa*-PolB2 and *Pa*-

PolB3 in an *in vitro* assay, because the heat-labile dRP residues are readily cleaved at the high incubation temperature required for their optimal activity; we could therefore not distinguish between non-enzymatic and enzymatic cleavage. Nevertheless, both enzymes yielded ligatable substrates in our *in vitro* assays (Fig. 5C), which attests to their ability to catalyze the removal of this blocking lesion. Moreover, the AP-site containing substrate was readily cleaved by *Pa*-PolB2 and/or *Pa*-DNA-Ligase by β-elimination (Fig. 5B, lane 3; data not shown), reflecting an intrinsic AP-lyase activity in these enzymes. It is therefore possible that the 5'-dRP group is removed by these enzymes in a BER reaction. A similar BER pathway was proposed to take place in mitochondria by Bogenhagen *et al.*, where both DNA polymerase γ and mtDNA ligase appear to possess a dRPlyase activity (43).

Based on our results, it is difficult to implicate either polymerase in *P. aerophilum* BER *in vivo*. Both *Pa*-PolB2 and *Pa*-PolB3 carried out the gap-filling reactions with similar efficiencies, and the yields of the repaired 60-mer products were comparable in assays using either enzyme. However, the enzymatic properties of the former enzyme resembled pol-β in many aspects: *Pa*-PolB2 contains several basic amino acid residues at its N-terminus that might impart it with a dRPase activity, it has limited processivity that is not stimulated by *Pa*-PCNA1 (Fig. 3D), and it is thus less prone to carry out strand displacement (Fig. 5C, cf. Lanes 10 and 11). In contrast, the primary structure and biochemial properties of crenarchaeal B3-DNA polymerases resemble more enzymes of the B1 family (40, 48-50); the B1 and B3 polymerases are therefore most likely involved in DNA replication in Crenarchaea (34). Thus, by analogy with the mammalian BER systems, where pol-β appears to be the major BER polymerase, but where the involvement of pol-δ in long patch BER could not be excluded (45), *Pa*-PolB2 would appear to be a better candidate for carrying out the gap-

filling function in short-patch BER in *P. aerophilum*, even though *Pa*-PolB3 may also participate in this process.

The role of *Pa*-PCNA1 in the *P. aerophilum* BER process *in vitro* is puzzling. This homotrimeric sliding clamp, the primary function of which is to increase the processivity of replicative DNA polymerases [reviewed in (51)], failed to stimulate *Pa*-PolB2. On the contrary, its presence in the reaction appeared to restrict the activity of *Pa*-PolB2 in the gap-filling reactions to the addition of a single nucleotide, while suppressing its strand displacement activity (Fig. 5). Yet, although *Pa*-PolB2 has no consensus PCNA binding motif, it appeared to interact with the sliding clamp in the EMSA assay on a substrate containing a single nucleotide gap, and the presence of *Pa*-PCNA1 in the reaction improved efficiency of the ligation reaction, especially at mispaired termini (Fig. 4C). It is possible that the association of the polymerase and the ligase with *Pa*-PCNA1 leads to the stabilization of DNA termini, which might otherwise fray at the elevated temperatures employed in these assays. *Pa*-PCNA1 might thus be fulfilling a role of a molecular matchmaker or scaffold protein similar to that postulated for XRCC1 in mammalian systems (19). It is interesting to note in this regard that the *P. aerophilum* genome does not appear to encode an XRCC1 homologue.

In conclusion, we have identified homologues of the mammalian BER proteins in the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum* and could show that these proteins can carry out efficient G•U $\rightarrow$ G•C repair in an oligonucleotide substrate. This is to our knowledge the first report describing the reconstitution of the archaeal BER process from purified recombinant proteins.

Acknowledgements- The authors wish to express their gratitude to Mahmud Shivji for the *P. aerophilum* extracts, to Samuel H. Wilson for the purified recombinant human DNA polymerase β and to Lingaraju Gondichatnahalli for the purified *Pa*-PCNA1. We thank Primo Schär for critical reading of the manuscript. The work was supported by a grant to A. S. from UBS AG.

REFERENCES

1. Woese, C. R., and Fox, G. E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5088-5090.
2. Keeling, P. J., and Doolittle, W. F. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 5761-5764.
3. Bell, S. D., Jaxel, C., Nadal, M., Kosa, P. F., and Jackson, S. P. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 15218-15222.
4. Cann, I. K., and Ishino, Y. (1999) *Genetics* **152**, 1249-1267.
5. Fitz-Gibbon, S. T., Ladner, H., Kim, U. J., Stetter, K. O., Simon, M. I., and Miller, J. H. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 984-989.
6. Skorvaga, M., Raven, N. D., and Margison, G. P. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 6711-6715.
7. Reilly, M. S., and Grogan, D. W. (2002) *FEMS Microbiol. Lett.* **208**, 29-34.
8. Friedberg, E. C. (2001) *Nat. Rev. Cancer* **1**, 22-33.
9. Yang, H., Fitz-Gibbon, S., Marcotte, E. M., Tai, J. H., Hyman, E. C., and Miller, J. H. (2000) *J. Bacteriol.* **182**, 1272-1279.
10. Sartori, A. A., Schar, P., Fitz-Gibbon, S., Miller, J. H., and Jiricny, J. (2001) *J. Biol. Chem.* **276**, 29979-29986.
11. Sartori, A. A., Fitz-Gibbon, S., Yang, H., Miller, J. H., and Jiricny, J. (2002) *EMBO J.* **21**, 3182-3191.
12. Lindahl, T. (1993) *Nature* **362**, 709-715.
13. Lindahl, T. (2001) *Prog. Nucleic Acid Res. Mol. Biol.* **68**, xvii-xxx
14. Scharer, O. D., and Jiricny, J. (2001) *Bioessays* **23**, 270-281.
15. Friedberg, E. C., Walker, G. C., and Siede, W. (1995) *DNA Repair and Mutagenesis*, American Society of Microbiology, ASM Press, Washington, D. C.
16. Dogliotti, E., Fortini, P., Pascucci, B., and Parlanti, E. (2001) *Prog. Nucleic Acid Res. Mol. Biol.* **68**, 3-27
17. Matsumoto, Y., and Kim, K. (1995) *Science* **269**, 699-702.
18. Vidal, A. E., Boiteux, S., Hickson, I. D., and Radicella, J. P. (2001) *EMBO J.* **20**, 6530-6539.
19. Kubota, Y., Nash, R. A., Klungland, A., Schar, P., Barnes, D. E., and Lindahl, T. (1996) *EMBO J.* **15**, 6662-6670.
20. Podlutsky, A. J., Dianova, II, Podust, V. N., Bohr, V. A., and Dianov, G. L. (2001) *EMBO J.* **20**, 1477-1482.
21. Klungland, A., and Lindahl, T. (1997) *EMBO J.* **16**, 3341-3348.
22. Gary, R., Kim, K., Cornelius, H. L., Park, M. S., and Matsumoto, Y. (1999) *J. Biol. Chem.* **274**, 4354-4363.
23. Yang, H., Chiang, J. H., Fitz-Gibbon, S., Lebel, M., Sartori, A. A., Jiricny, J., Slupska, M. M., and Miller, J. H. (2002) *J. Biol. Chem.* **277**, 22271-22278.
24. Hudepohl, U., Reiter, W. D., and Zillig, W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5851-5855.
25. Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248-254.
26. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
27. Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387-395.
28. Mol, C. D., Hosfield, D. J., and Tainer, J. A. (2000) *Mutat. Res.* **460**, 211-229.
29. Hosfield, D. J., Guan, Y., Haas, B. J., Cunningham, R. P., and Tainer, J. A. (1999) *Cell* **98**, 397-408.
30. Sobol, R. W., Horton, J. K., Kuhn, R., Gu, H., Singhal, R. K., Prasad, R., Rajewsky, K., and Wilson, S. H. (1996) *Nature* **379**, 183-186.
31. Srivastava, D. K., Berg, B. J., Prasad, R., Molina, J. T., Beard, W. A., Tomkinson, A. E., and Wilson, S. H. (1998) *J. Biol. Chem.* **273**, 21203-21209.
32. Perler, F. B., Kumar, S., and Kong, H. (1996) *Adv. Protein Chem.* **48**, 377-435
33. Edgell, D. R., Klenk, H. P., and Doolittle, W. F. (1997) *J. Bacteriol.* **179**, 2632-2640.
34. Bohlke, K., Pisani, F. M., Rossi, M., and Antranikian, G. (2002) *Extremophiles* **6**, 1-14.
35. Lindahl, T., and Barnes, D. E. (1992) *Annu. Rev. Biochem.* **61**, 251-281
36. Timson, D. J., Singleton, M. R., and Wigley, D. B. (2000) *Mutat. Res.* **460**, 301-318.
37. Martin, I. V., and MacNeill, S. A. (2002) *Genome Biol.* **3**
38. Haas, B. J., Sandigursky, M., Tainer, J. A., Franklin, W. A., and Cunningham, R. P. (1999) *J. Bacteriol.* **181**, 2834-2839.
39. Levin, J. D., Shapiro, R., and Demple, B. (1991) *J. Biol. Chem.* **266**, 22893-22898.
40. Kahler, M., and Antranikian, G. (2000) *J. Bacteriol.* **182**, 655-663.

41. Wiebauer, K., and Jiricny, J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5842-5845
42. Goffin, C., Bailly, V., and Verly, W. G. (1987) *Nucleic Acids Res.* **15**, 8755-8771.
43. Bogenhagen, D. F., Pinz, K. G., and Perez-Jannotti, R. M. (2001) *Prog. Nucleic Acid Res. Mol. Biol.* **68**, 257-271
44. Prasad, R., Beard, W. A., Chyan, J. Y., Maciejewski, M. W., Mullen, G. P., and Wilson, S. H. (1998) *J. Biol. Chem.* **273**, 11121-11126.
45. Krokan, H. E., Drablos, F., and Slupphaug, G. (2002) *Oncogene* **21**, 8935-8948.
46. Yang, H., Phan, I. T., Fitz-Gibbon, S., Shivji, M. K., Wood, R. D., Clendenin, W. M., Hyman, E. C., and Miller, J. H. (2001) *Nucleic Acids Res.* **29**, 604-613.
47. Sobol, R. W., Prasad, R., Evenski, A., Baker, A., Yang, X. P., Horton, J. K., and Wilson, S. H. (2000) *Nature* **405**, 807-810.
48. Cann, I. K., Ishino, S., Nomura, N., Sako, Y., and Ishino, Y. (1999) *J. Bacteriol.* **181**, 5984-5992.
49. Daimon, K., Kawarabayasi, Y., Kikuchi, H., Sako, Y., and Ishino, Y. (2002) *J. Bacteriol.* **184**, 687-694.
50. Fogg, M. J., Pearl, L. H., and Connolly, B. A. (2002) *Nat. Struct. Biol.* **9**, 922-927
51. Kelman, Z. (1997) *Oncogene* **14**, 629-640
52. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673-4680.
53. Brautigam, C. A., and Steitz, T. A. (1998) *Curr. Opin. Struct. Biol.* **8**, 54-63.
54. Blanco, L., and Salas, M. (1995) *Methods Enzymol.* **262**, 283-294
55. Shamoo, Y., and Steitz, T. A. (1999) *Cell* **99**, 155-166.

FIGURE LEGENDS

Fig. 1. BER activity in *P. aerophilum* whole cell-free extracts and putative BER proteins identified in the genomic sequence of this organism.

A) Base excision repair activities in whole cell-free extracts (WCE) of *P. aerophilum*. 1 pmol of the 5'-end-labeled substrates (schematically drawn above the figure), 1nt-gap (lanes 1-6) and G•U mismatch (lanes 7-12), were incubated with 20 μg of *P. aerophilum* WCE for 30 min at 60°C. Reactions were supplemented with 5mM $MgCl_2$, 20μM dNTPs and 2mM ATP as indicated. Lanes 1 and 7, no WCE was added. To ensure complete ligation of the BER intermediates, pre-adenylated T4 DNA Ligase was added and incubation was allowed to proceed for a further 1 hour at 30°C. The positions of the reaction products are indicated on the right. The 60-mer band in lane 8 is due to incomplete processing of the uracil-containing oligonucleotide by the extract. **B)** *P. aerophilum* AP endonuclease belongs to the AP endonuclease IV family. Alignment of the complete amino acid sequence of *P. aerophilum* AP endonuclease IV (*Pa*-EndoIV) with that of its *E. coli* homologue (*Ec*-EndoIV; Swiss-Prot: P12638). Identical residues are boxed and shaded in light gray. The two sequences were aligned using the CLUSTAL W program (52). The nine metal-binding active site residues are marked by an arrowhead. **C)** Partial sequence alignments of *P. aerophilum* DNA polymerases B1, B2 and B3 with the gene 43 protein (gp43) of bacteriophage RB69, a prototype of the DNA polymerase B family (Swiss-Prot: Q38087) (53). Only amino acid residues belonging to the conserved regions of family B DNA polymerases are shown, and identical residues are shaded in light gray. The three highly-conserved motifs Pol-I through Pol-III, indicative of family B DNA polymerases, were identified in all three *P. aerophilum* DNA polymerases, whereas motifs forming the conserved 3→5 exonuclease active site (Exo

I, Exo II and Exo III), were only present in *Pa*-PolB3 and *Pa*-PolB1. A single putative 3→5 exonuclease motif, Exo III, was also identified in *Pa*-PolB2. Amino acid residues that have been identified as functionally-important by mutational studies are marked by arrowheads (54). The two invariant aspartic acid residues forming the catalytic dyad of all known polymerases in family B are marked with an asterisk. The DNA-binding motif YxGA, which plays a critical role in the crosstalk between synthesis and degradation, is indicated by a bar. Putative sliding clamp-binding motifs at the C-terminus of *Pa*-PolB3 and *Pa*-PolB1 are aligned with the consensus amino acid sequence of the clamp-binding peptide in gp43 (LFDMF) (55). **D)** The *P. aerophilum* genome sequence encodes an ATP-dependent DNA ligase. *Upper panel,* domain structure of ATP-dependent ligases from *P. aerophilum* and *H. sapiens*. The 589-amino acid *P. aerophilum* DNA ligase (*Pa*-DNA-Ligase) and the 919-amino acid *H. sapiens* DNA ligase I (*Hs*-DNA-Ligase1) polypeptides are depicted as *straight lines* with the positions of the conserved motifs I, III, IIIa, IV, V, and VI denoted by boxes. Abbreviations: CD, catalytic domain; NCD, conserved non-catalytic domain of unknown function. The nuclear form of DNA ligase I found in vertebrates and yeast has an additional N-terminal extension bearing the PCNA binding motif (PBM) and the nuclear localization signal (NLS). *Lower panel,* partial sequence alignment of the catalytic domain (CD) of *Pa*-DNA Ligase with *Hs*-DNA-Ligase1 (Swiss-Prot: P18858). Boxes I to VI represent the six motifs commonly found in ATP-dependent DNA ligases. Identical residues are shaded and the lysine nucleophile in Motif I (KxDGxR), to which AMP becomes covalently linked during ligase-adenylate formation, is marked by an arrowhead. **E)** Purified recombinant proteins from *P. aerophilum* used in this study. The proteins were purified as described under "Experimental Procedures", electrophoresed in 10% SDS-PAGE gels and

visualized by Coomassie Blue staining. M, molecular mass standards (kDa).

Fig. 2. AP endonuclease activity of *Pa*-EndoIV.

A) The AP endonuclease activity of recombinant *Pa*-EndoIV is thermostable. AP-sites were processed by hot alkali treatment (0.1M NaOH for 10 min at 90°C, lane 2) or incubated with soluble (sol) or partially-purified (Ni-NTA) fractions obtained from *E. coli* BL21 cells transfected with the empty pET28c(+) vector (lanes 3, 4 and 7) or with the pET28-*PaEndoIV* plasmid (lanes 5, 6, 8 and 9). 1 pmol of the G•AP* substrate was incubated for 30 min at 37°C with 2 µl of the fractions as described under "Experimental Procedures". HT indicates that the proteins were heat-treated for 15 min at 70°C prior to addition of the substrate (lanes 4, 6 and 9). This experiment shows that no contaminant *E. coli* AP-site processing activity was present after Ni-NTA purification. The bands due to the cleaved 23-mer fragment and to the 60-mer substrate are indicated on the right. **B)** *Pa*-EndoIV activity increases at higher temperatures and resists inactivation by EDTA. 50 nmol of the G•AP* substrate were incubated either without or with 5 nmol of highly-purified recombinant *Pa*-EndoIV for 30 min at 25°C (lanes 1 and 2), 37°C (lanes 3 and 4), 50°C (lanes 5 and 6) and 60°C (lanes 7 and 8) in the presence of 5mM EDTA. **C)** Combined uracil-DNA glycosylase and AP endonuclease activities of recombinant *P. aerophilum* proteins. 1 pmol of labeled 60-mer G•U* substrate was incubated with 0.1 pmol of *Pa*-EndoIV alone (lane 1), or with 0.1 pmol *Pa*-UDGa or 1 pmol *Pa*-UDGb (lanes 2 and 3 resp.), two monofunctional uracil-DNA glycosylases from *P. aerophilum*. Incubation was carried out for 15 min at 60°C. AP-site cleavage by NaOH treatment converts the 60-mer substrate to a 23-mer product with a 3'-phosphate group, which migrates slightly faster through the gel (lanes 4 and 5) than the product generated by AP endonuclease cleavage, which generates free 3'-hydroxyl termini that are substrates for a DNA polymerase (lanes 6 and 7). **D)** Combined uracil-DNA

glycosylase and AP endonuclease activities in *P. aerophilum* WCE. 1 pmol of the labeled G•U* (lanes 3-6) or G•C* (lane 7) substrates was incubated for 30 min at 60° with 5 µg of *P. aerophilum* WCE. Lane 1, no proteins added; lane 2, purified recombinant *Pa*-UDGa and *Pa*-EndoIV. The WCE of P. aerophilum excises the uracil from a G•U mismatch and cleaves the resulting AP-site. This combined action does not appear to require Mg^{2+} as a cofactor (lane 3) and the responsible proteins in the WCE are heat-resistant (lane 4; WCE was pre-treated for 10 min at 85°C). Upon addition of 100µM dNTPs, the Mg^{2+}-dependent incorporation of several nucleotides was observed, reflecting the activity of a DNA polymerase(s) in the extract (lanes 5 and 6).

Fig. 3. Biochemical characterization of DNA polymerase B2 of *P. aerophilum* (*Pa*-PolB2). **A)** DNA synthesis by *Pa*-PolB2 is enhanced at higher temperatures. The primer- extension ability of partially-purified Ni-NTA column fractions obtained from *E. coli* BL21 cells transfected either with the empty pET28c(+) vector (lane 5) or with the pET28-*PaPolB2* plasmid (lanes 6-8) were compared with 1 unit of either T4 DNA polymerase (lane 2) or *Taq* polymerase (lanes 3 and 4, respectively). Incubation was performed for 15 min at the indicated temperatures. The bands due to the 23-mer primer and the fully-extended 60-mer reaction product are indicated on the right. This experiment shows that no contaminating *E. coli* DNA polymerase was present in these fractions after Ni-NTA purification. **B)** *Pa*-PolB2 lacks a 3'→5 exonuclease activity and is aphidicolin-resistant. Single- and double-strand-dependent 3→5 exonuclease activities and the effect of aphidicolin on DNA polymerases from *P. aerophilum* (*Pa*-PolB2 and *Pa*-PolB3) were compared with DNA polymerases from *T. aquaticus* (*Taq* pol) and *P. furiosus* (*Pfu* pol). Reactions were carried out for 1 h at 60°C,

using 0.25 units of each *Taq* pol (lanes 1-3) and *Pfu* pol (lanes 4-8) and 1 pmol of each *Pa*-PolB2 (lanes 9-13) and *Pa*-PolB3 (lanes 14-18). In the inhibition experiments, the DNA polymerases were pre-incubated with 2mM aphidicolin (A) or 10% DMSO (D) for 5 min at room temperature. In all reactions, *Pa*-PolB3 was additionally heat-treated for 15 min at 80°C prior to use. The sizes of the non-degraded primer and the reaction products are indicated. **C)** *Pa*-PolB2 and human pol-β behave similarly in inhibition assays. The effects of 2mM aphidicolin (A), 10% DMSO (D) and 1mM ddCTP (dd) on *Pa*-PolB2 were compared with human DNA polymerase β (pol-β) and *Pfu* pol. Lane 1, no enzymes added. 1 pmol of pol-β (lanes 2-5), 0.25 units of *Pfu* pol (lanes 6-9) and 1 pmol of *Pa*-PolB2 (lanes 10-13) were incubated with 1 pmol of the primer-extension template for 10 min (at 37°C or 60 °C) in the presence of 100μM dNTPs. The bands due to the 23-mer primer substrate and the 60-mer product are indicated. The bands resulting from the incorporation of ddCMP are marked with asterisks. **D)** DNA synthesis by *Pa*-PolB3, but not *Pa*-PolB2, is stimulated by the recombinant *P. aerophilum* sliding clamp protein *Pa*-PCNA1. Under standard primer extension conditions, 0.5 pmol of *Pa*-PolB2 (lanes 2-5) or 0.1 pmol of *Pa*-PolB3 (lanes 6-9) was incubated for 5 and 15 min at 65°C. All reactions contained 1 pmol of labeled primer-extension substrate and 5 pmol of unlabeled poly(dA):oligo(dT) substrate (see DNA ligase assay) as competitor. In lanes 4, 5, 8 and 9, *Pa*-PCNA1 (500 ng) was included in the reaction mixtures. **E)** *Pa*-PolB2, but not *Pa*-PolB3, specifically binds to a 1nt-gap and the binding is stimulated by *Pa*-PCNA1**.** *P. aerophilum* proteins *Pa*-UDGb (5 pmol), *Pa*-PolB2 (1 or 2 pmol) and *Pa*-PolB3 (2 pmol) were incubated with the G•C* homoduplex (lanes 1-4) or the 1nt-gap (lanes 5-13, illustrated in Figure 1A) substrates for 30 min at 4°C. Lanes 1 and 5, no proteins added. Lanes 2 and 6, *Pa*-UDGb. A stable polymerase/DNA complex was formed

only between *Pa*-PolB2 and the substrate containing a 1 nt-gap (lane 9, complex 2). This complex was slightly supershifted and increased in intensity in the presence of 500 ng of *Pa*-PCNA1 (lanes 10 and 11, complex 3). The bands corresponding to the free probe and the different protein/DNA complexes are indicated on the right.

Fig. 4. Characterization of the *P. aerophilum* DNA-Ligase (*Pa*-DNA-Ligase).

A) *Pa*-DNA-Ligase forms a covalent enzyme-AMP (E-AMP) complex in the presence of ATP as co-factor. 5 pmol of *Pa*-DNA-Ligase (~ 64 kDa) and 40 units of T4 DNA Ligase (56 kDa) were incubated with [α-^{32}P]ATP for 15 min at the indicated temperatures. The reactions were stopped by boiling in SDS-PAGE loading buffer and analyzed on a 10% SDS-polyacrylamide gel. **B)** *Pa*-DNA-Ligase efficiently ligates single-strand breaks *in vitro*. 0.5 pmol of the labeled substrate was incubated with increasing amounts of the recombinant *Pa*-DNA-Ligase (0.1, 1, and 10 pmol) in 1 x T4 DNA Ligase buffer for 15 min at 55°C. The ligation efficiency was compared with that mediated by T4 DNA ligase at 37°C (~ 6 Weiss units) and monitored by converting the oligo-dT (24-mer) into higher order oligomers (48-, 72-, 96-, 120-mers) as indicated. **C)** Nick-joining activity of *Pa*-DNA-Ligase is reduced in *P. aerophilum* whole cell-free extracts (WCE). Ligation reactions were carried out for 1 h at 60°C by incubating *Pa*-DNA-Ligase (5 pmol) and WCE (20 µg) with substrates containing either a G•C base pair (lanes 1-8) or a G•T mismatch (lanes 9-16) at the 3'-end of the 23-mer fragment. Lanes 1 and 9, no proteins added; lanes 2 and 10, 40 units of T4 DNA Ligase; lanes 3-5 and 11-13, 5 pmol of *Pa*-DNA-Ligase; lanes 5 and 13, 500 ng of *Pa*-PCNA1; lanes 6-8 and 14-16, 20 µg of *P. aerophilum* WCE. The bands due to the nicked substrates and the ligated products are indicated.

Fig. 5. *In vitro* reconstitution of BER using recombinant *P. aerophilum* proteins.

A) *In vitro* repair of a single-strand break by recombinant *P. aerophilum* enzymes is affected by *Pa*-PCNA1. The gap-filling properties of *Pa*-PolB2 (0.5 pmol B2, even-numbered lanes) and *Pa*-PolB3 (0.1 pmol B3, odd-numbered lanes), resulting in the formation of ligatable BER intermediates, were compared by using either dCTP alone (lanes 1-7) or all four dNTPs (lanes 8-15). The 1-nt gap substrate (Figure 1A) was incubated for 30 min at 60°C in the presence or absence of *Pa*-DNA-Ligase and *Pa*-PCNA1 as indicated. Lane 1, no enzymes added. The bands due to the nicked 23-mer substrate, the 24-mer intermediate resulting from the extension of the primer by a single nucleotide (dCMP) and the 60-mer product are indicated. **B)** AP-site repair by a concerted action of *Pa*-EndoIV, *Pa*-PolB2 and *Pa*-DNA-Ligase. The repair of normal (AP) and reduced abasic (RAP) sites was examined by using the minimum repertoire of enzymes needed to cleave the AP-site, incorporate the missing nucleotide (dCMP) and religate the nick. 1 pmol of the substrate, G•AP* or G•RAP* (illustrated on the right), was incubated with *Pa*-EndoIV (0.2 pmol), *Pa*-PolB2 (0.5 pmol) and *Pa*-DNA-Ligase (5 pmol) for 30 min at 60 °C. The reactions were initiated by the addition of either 20μM dCTP alone (lanes 1-6) or 20μM each of dCTP, dATP and dGTP (lanes 7 and 8). Lane 1, incubation of the AP substrate without enzymes at 37°C; lane 2, incubation of the AP substrate without enzymes at 60°C; lane 3, no *Pa*-EndoIV added; lane 4, no *Pa*-PolB2 added; lanes 5 and 7, no *Pa*-DNA-Ligase added; lanes 6 and 8, all three proteins added. The bands due to the endonucleolytically-cleaved AP-site (23-mer) and the different reaction products are indicated. **C)** Reconstitution of DNA base excision-repair with *P. aerophilum* proteins. The fluorescently-labeled substrate DNA containing a single G•U mismatch was incubated with the recombinant *P. aerophilum* enzymes as indicated. The

amounts of proteins used were: 1 pmol *Pa*-UDGb, 0.2 pmol *Pa*-EndoIV, 500 ng *Pa*-PCNA1, 5 pmol *Pa*-DNA-Ligase and 0.5 pmol of the DNA polymerases (*Pa*-PolB2 and *Pa*-PolB3). The bands due to the reaction products and intermediates are indicated. The band migrating at 50 nucleotides results from a DNA polymerase-catalyzed strand displacement of the oligonucleotide downstream from the pre-incised AP-site.

Table I
Substrate oligonucleotide sequences (5‘-3‘)

60-mer G	TAGACATTGCCCTCGAGGTACCATGGATCCGATGTCGACCTCAAACCTAGACGAATTCCG
60-mer U-F	CGGAATTCGTCTAGGTTTGAGGTUGACATCGGATCCATGGTACCTCGAGGGCAATGTCTA
60-mer C-F	CGGAATTCGTCTAGGTTTGAGGTCGACATCGGATCCATGGTACCTCGAGGGCAATGTCTA
23-mer-F	CGGAATTCGTCTAGGTTTGAGGT-**OH**
36-mer	**P**-GACATCGGATCCATGGTACCTCGAGGGCAATGTCTA
24-mer C-F	CGGAATTCGTCTAGGTTTGAGGTC-**OH**
24-mer T-F	CGGAATTCGTCTAGGTTTGAGGTT-**OH**
50-mer G	CCTCGAGGTACCATGGATCCGATGTCGACCTCAAACCTAGACGAATTCCG

F, denotes 5‘-labeled with Fluorescein.

Figure 1, Sartori and Jiricny

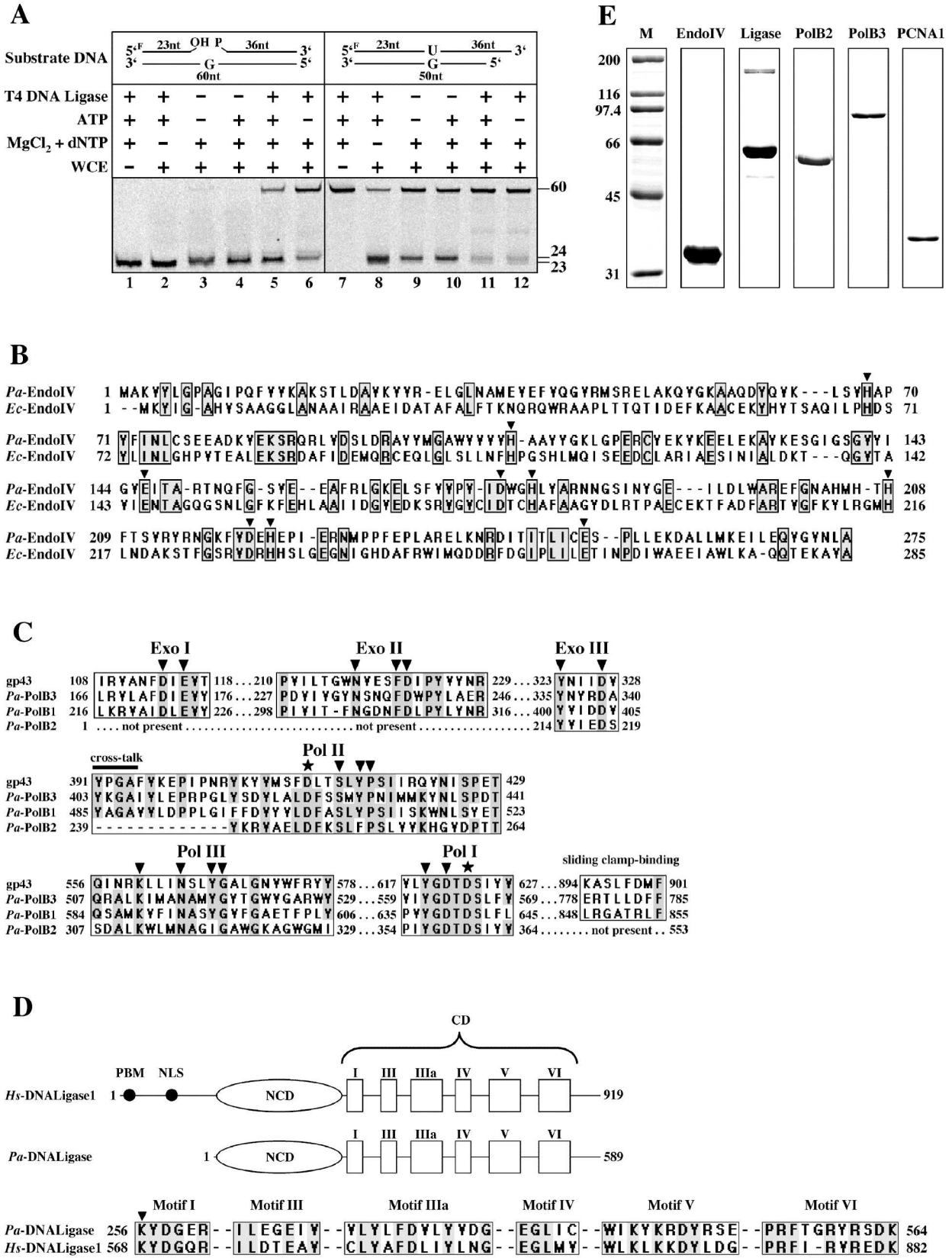


Figure 2, Sartori and Jiricny

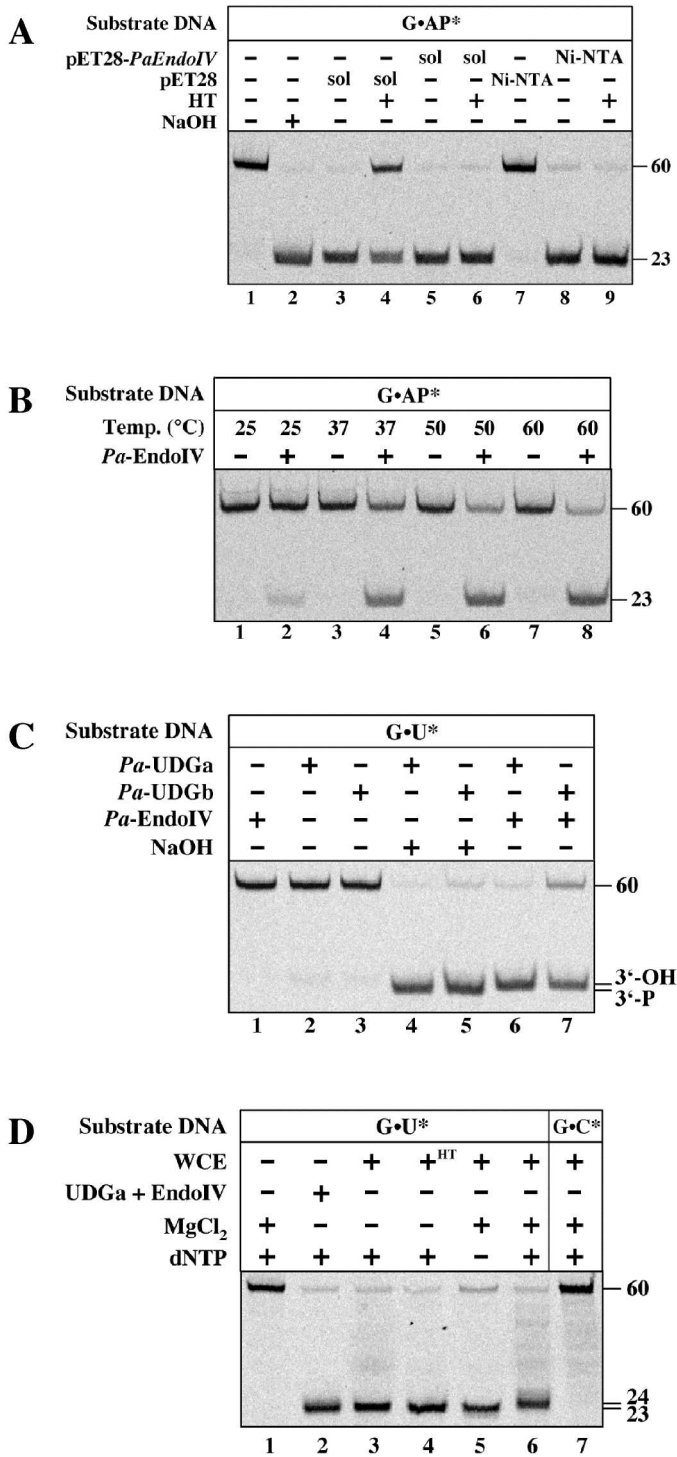


Figure 3, Sartori and Jiricny

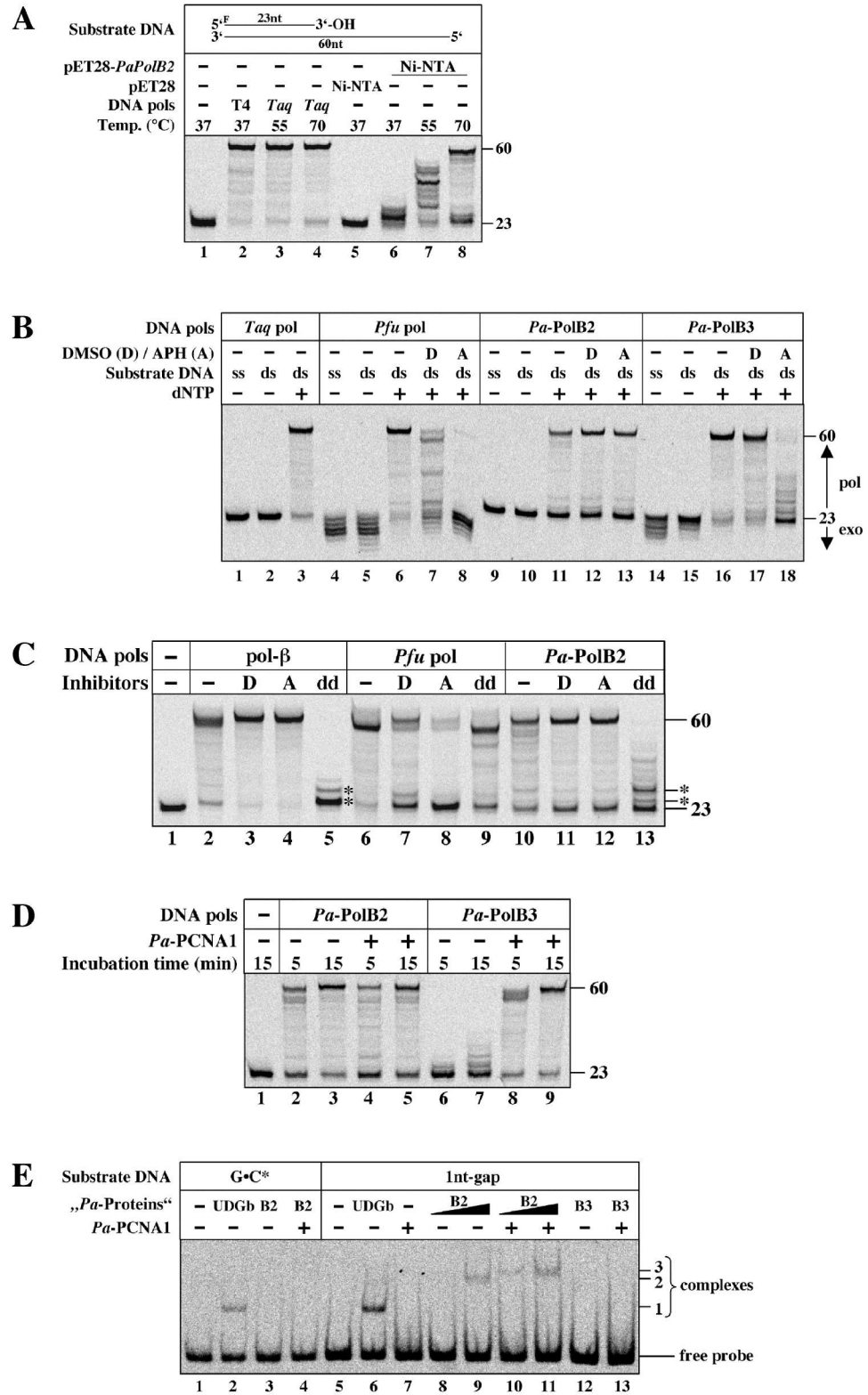
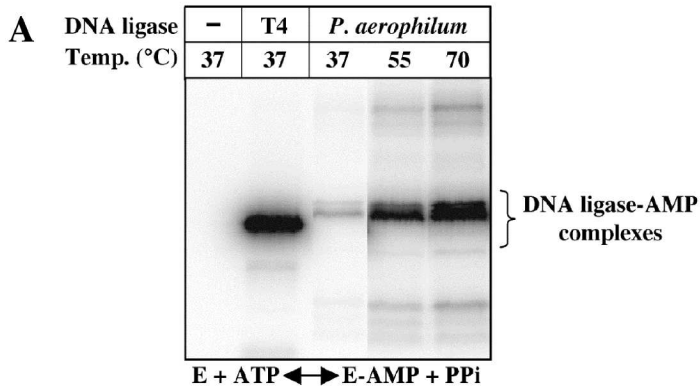
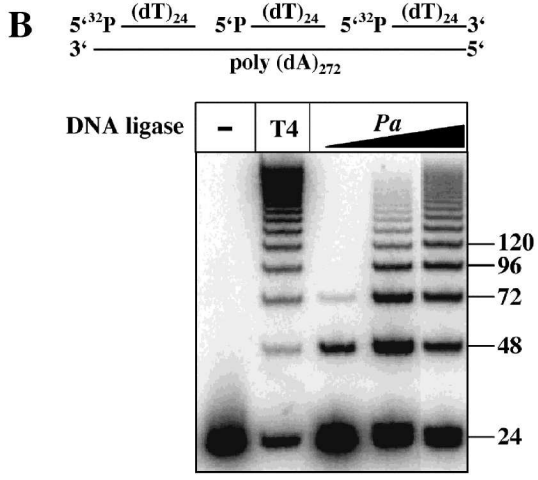


Figure 4, Sartori and Jiricny





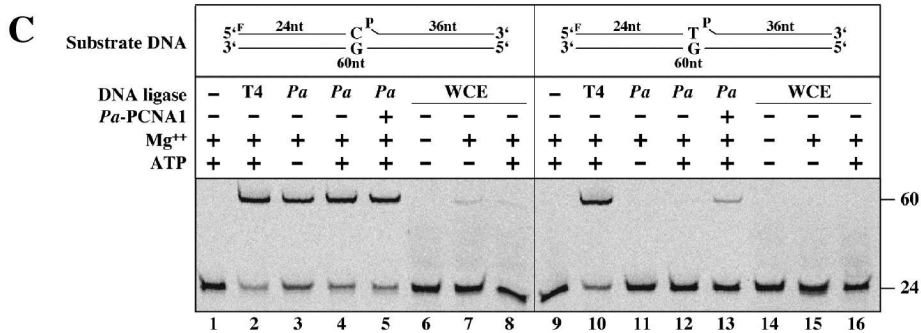


Figure 5, Sartori and Jiricny